94. (Amended) The nucleic acid of claim 57, wherein the [functional] additional domain is heterologous with respect to the two nucleic acid-binding domains.

96. (Amended) The method of claim 76, wherein the [functional] additional domain is heterologous with respect to the two nucleic acid-binding domains.

Remarks

Claims 40-70 and 71-88 are pending. Claim 40 has been amended to correct for typographical errors. Claims 57, 58, 63, 65, 76, 77, 94 and 96 have been amended for improved clarity and to be in better condition for appeal. No new matter has been added.

Cancellation and or amendment of claims should in no way be construed as an acquiescence to any of the Examiner's rejections. The cancellation and/or amendments to the claims are being made solely to expedite prosecution of the present application. Applicants reserve the option to further prosecute the same or similar claims in the instant or in a subsequent patent application.

Rejection of claims 40-70, 72, 89-92, 94-95, and 97 under 37 C.F.R. § 103 (a) in view of Park et al., Mitchell et al., Harrison and Schultz

Claims 40-70, 72, 89-92, 94-95, and 97 have been rejected under 37 C.F.R. § 103 (a) as being unpatentable over Park et al. (*PNAS* 89: 9094 (1992)), in view of Mitchell et al. (*Science* 245:371 (1989)), Harrison (*Nature* 353:715 (1991)), and Schultz (*Nature* 240: 426 (1988)). Applicants respectfully traverse this rejection.

Claim 40, and claims 41-65 and 72 dependent therefrom, are drawn to a nucleic acid encoding a chimeric protein which binds to a nucleic acid comprising a composite binding site, the chimeric protein comprising two nucleic acid-binding domains, each of which binds a sequence which is a portion of the composite binding site, and wherein only one of the two nucleic acid-binding domains includes a zinc finger motif. Claim 66, and claims 67-70 dependent therefrom, are drawn to a nucleic acid encoding a chimeric protein which binds to a nucleic acid comprising a composite binding site, the chimeric protein comprising two nucleic acid-binding domains, each of which binds a sequence which is a portion of the composite binding site, and wherein only one of the two nucleic acid-binding domains is a nucleic acid-binding domain from a homeodomain containing protein.

Park et al. is relied on by the Examiner as teaching "a general strategy for designing proteins to recognize specific DNA-binding sites" and that "[t]his technique creates a protein that recognizes the composite site (page 9094, column 1)." The Examiner indicates that Park et al. "do not teach to specifically use the DNA-binding domains from distinct families of nucleic acid binding domains, use of specific types of domains such as zinc-finger domains."

Mitchell et al. is relied on by the Examiner as teaching that "different DNA binding transcription factors are composed of a surprising variety of usually separable DNA binding and transcriptional activation domains (page 372, column 2)."

Harrison is relied on as teaching that "many DNA-binding proteins recognize specific sites through small, discrete domains and that these domains can be interchanged between proteins, showing that these domains are independent folded units."

Schultz is relied on as teaching that "enzymes can be created by adding or replacing entire binding or catalytic domains to generate hybrid enzymes with novel specificities" and that "[s]elective fusion of nucleic acid-specific binding domains may produce sequence-specific DNA or RNA cleaving enzymes (page 431, column 1)."

It is the Examiner's position that

> [i]t would have been obvious to one of skill in the art at the time the invention was made to use the various DNA binding domains, activation domains, and cleavage domains taught by Mitchell et al., Harrison, and Schulz in the general strategy for designing proteins to recognize specific DNA-binding sites taught by Park et al. because Park et al. teach that it is within the ordinary skill in the art to stitch the DNA binding domains together from any proteins that recognize a specific DNA sequence by binding along the major groove, to recognize a composite site and Mitchell et al., Harrison, and Schultz teach such domains that can be functionally separated and recombined with other domains.

Applicants submit that the cited references alone, or together with the general knowledge in the art at the time the invention was filed, fail to provide neither sufficient motivation to combine them to obtain the claimed nucleic acids nor the requisite reasonable expectation of success.

However, applicants did not simply infer, as asserted in the Office Action, that "because one of the references in the rejection did not teach the whole invention, there would not have been a reasonable expectation of success." On the contrary, Applicants' noted that Park et al. fail to provide any motivation to combine the references and to provide a reasonable expectation of success, and that none of the secondary references cited cured this defect.

The heart of the Examiner's position appears to be that Park et al. alone "provides very strong motivation to make chimeric DNA-binding proteins that bind to composite sequences by fusing two previously separate DNA binding domains together." Applicants disagree.

Applicants respectfully submit that the Examiner's conclusion requires one to read the excerpted Park et al sentences out of context. In view of the actual disclosure of the Park et al reference, a person of skill in the art would understand the general statements made in Park et al. to refer to arms of dimeric DNA binding proteins, and would not give broader meaning to statements in that reference that could otherwise be given broader interpretation, i.e., if taken out of that context.

There are several lines of evidence that even Park et al. did not intend the general statements in their paper to be given the broader interpretation reflected in the office action.

First, Park et al.'s later published paper (*PNAS* 90: 4892 (1993); copy of which is enclosed herein as Exhibit A), defines the concept of protein stitchery, referring to their '92 paper, as representing that "individual basic arms (half sites) of the dimer and the individual half sites of the DNA can be recombined or stitched together in various sequences to form new proteins selective for binding to the new DNA sites" (emphasis added; see paragraph bridging pages 4892 and 4893). Thus, Park et al. defined the general concept of protein stitchery as combining "arms" of proteins which normally form dimers.

Second, if Park et al truly intended the generic meaning represented by the Examiner, they wouldn't have limited themselves to proteins which bind along the major groove of the DNA. Clearly there are DNA binding proteins which bind along the minor groove and indeed homeodomains and zinc finger domains, such as are used to illustrate the subject invention, do make minor groove contacts.

In addition, the publication of applicants' work in *Science*, as discussed further below, belies acceptance in the art of broader conclusions from the Park et al reference.

Accordingly, Applicants again submit that the general statements made by Park et al. would not have been viewed, and should not now be viewed, as being as broad as the Examiner contends, and would not provide the requisite motivation to make the claimed chimeric proteins.

Moreover, as Applicants explained before, Parks et al. cross-linked together the DNA binding domains of two proteins which normally bind DNA only in the form of a homodimer. However, Park et al. does not teach or suggest that a chimeric protein having a composite DNA binding domain consisting of two or more DNA binding domains from different types of DNA binding proteins, which do not normally interact with each other, would bind DNA with higher affinity to the composite binding site than to portions of it. Harrison and Mitchell et al. merely teach that DNA binding domains can be separated from the rest of DNA binding proteins. However, these teachings

do not fail to cure the defect of Park et al. Furthermore, neither Harrison nor Mitchell refer to portions of DNA binding domains, e.g, individual zinc finger domains, which can also constitute part of Applicants' claimed composite DNA binding domain.

The Examiner also states that "[a]bsent evidence to the contrary, there would have been a reasonable expectation of success that the domains taught by Mitchell et al. and Harrison could be combined with each other to create a protein that recognizes a composite binding site as taught by Park et al." However, as set forth above, Mitchell and Harrison merely teach that DNA binding domains can be separated from the rest of DNA binding proteins. There was no reasonable expectation of success that a chimeric protein containing a composite DNA binding domain consisting of DNA binding domains from proteins which are unrelated, as claimed by Applicants, would bind to a composite DNA binding site with higher affinity than to each of the half sites to which each of the DNA binding domains of the composite DNA binding domain bind. Nor was there any reasonable expectation of success that binding of such a chimeric protein containing a transcriptional activation domain to a DNA binding site would be able to stimulate transcription of a target gene operably linked to the DNA binding site, as Applicants showed.

Further support that there was no reasonable expectation of success to obtain the claimed invention is provided by a statements made by one of the inventors of the instant application after publication of their invention in the journal Science: "laboratory tests have proved the artificial switch can find, and control, a single gene among the 80,000 that exits in humans." The article (attached hereto as Exhibit B) also quotes Carl Pabo, as stating "the critical thing was showing it can bind the proper site." Furthermore, if there had been a reasonably expectation of success, Applicants' description of the invention would not have been published in the prestigious peer-reviewed *Science* journal (Pomerantz et al. (1995) *Science* 267: 93, attached hereto as Exhibit C).

With regard to making a nucleic acid and vector comprising the nucleic acid which encodes the chimeric protein, the Examiner states that "it would have been obvious to do so because Parks et al. teach that a continuous approximately 70 amino-acid protein that should recognize a predictable site can be made, instead of using a cysteine linker, and thus it would have been obvious to make a nucleic acid that encodes this protein and place the nucleic acid in a vector to express the protein, because such a way of making a mutated, recombinant protein is and was well known in the art." Applicants respectfully traverse this statement. Although Parks et al. may make general statements, i.e., speculate, that the cross-linking between the two monomer DNA binding domains (presumably of a dimeric DNA binding protein) could be replaced by a peptide bond, a person of skill in the art at the time the invention was made would have known that a peptidic bond may give rise to a protein structure that is different from that resulting from disulfide cross-linking. Thus, there was no reasonable expectation of success that two different DNA binding domains linked together through a peptidic bond would form a chimeric DNA binding protein.

Thus, in view of all of the above, Applicants respectfully request that the Examiner reconsider and withdraw rejection of claims 40-72 under 37 C.F.R. § 103 (a) as being unpatentable over Park et al., in view of Mitchell et al., Harrison, and Schultz.

Rejection of claims 40-70 and 72-88 under 37 C.F.R. § 103 (a) in view of Park et al., Mitchell et al., Harrison, Schultz and Gossen et al.

Claims 40-70 and 72-88 have been rejected under 37 C.F.R. § 103 (a) as being unpatentable over Park et al. (supra), in view of Mitchell et al. (supra), Harrison (supra), Schultz (supra) as applied to claims 40-70, 72, 89-92, 94-95, and 97 above, and further in view of Gossen et al. (U.S. Patent No. 5, 464,758). Applicants respectfully traverse this rejection.

Claim 40 and claims 41-65, and 72-74 dependent therefrom are drawn to a nucleic acid encoding a chimeric protein which binds to a nucleic acid comprising a composite binding site, the chimeric protein comprises two nucleic acid-binding domains, each of which binds a sequence which is a portion of the composite binding site, and wherein only one of the two nucleic acid-binding domains includes a zinc finger motif. Claim 66 and claims 67-70 dependent therefrom are drawn to a nucleic acid encoding a chimeric protein which binds to a nucleic acid comprising a composite binding site, the chimeric protein comprises two nucleic acid-binding domains, each of which binds a sequence which is a portion of the composite binding site, and wherein only one of the two nucleic acid-binding domains is a nucleic acid-binding domain from a homeodomain containing protein. Claims 75-83 and 84-89 are drawn to a method for modulating expression of a gene in a cell, comprising modulating the level of a chimeric protein in a cell which includes a gene operably to a composite binding site to which the chimeric protein binds, wherein the chimeric protein comprises two nucleic acid-binding domains, each of which binds a sequence which is a portion of the composite binding site

Park et al., Mitchell et al., Harrison and Schultz are relied upon by the Examiner as disclosing what is summarized in the previous section. Gossen et al. is relied upon by the Examiner as teaching "a nucleotide molecule coding for a chimeric transactivator fusion protein comprising a DNA binding domain (tet repressor binding domain) and a transactivation domain (such as VP16 of HSV)."

It is the Examiner's opinion that "[i]t would have been obvious to one of ordinary skill in the art at the time the invention was made to form a transcriptional regulatory system from the DNA encoding a chimeric transactivation protein made obvious by the teachings of Park et al. (AW2), Mitchell et al. (S), Harrison (T) and Schultz (U), using the method taught by Gossen et al. because Gossen et al. teach that it is within the ordinary skill in the art to make a nucleic acid vector that

encodes a chimeric transactivator fusion protein (under the control of a promoter active in eukaryotic cells), make a nucleic acid encoding a heterologous protein operably linked to a regulator binding site that the chimeric protein binds to..."

Applicants respectfully submit that, as set forth above, Park et al. (AW2), Mitchell et al. (S), Harrison (T) and Schultz (U) do not make obvious a chimeric transactivation protein. Thus, even if, as contended by the Examiner, Gossen et al. teach that "it is within the skill in the art to make a nucleic acid vector that encodes a chimeric transactivator fusion protein (...), make a nucleic acid encoding a heterologous protein operably linked to a regulator binding site that the chimeric proteins binds to, place the nucleic acid in a eukaryotic cell..." the cited references fail to provide any motivation to combine the references and to provide a reasonable expectation of success.

Thus, in view of all of the above, Applicants respectfully request that the Examiner reconsider and withdraw the rejection of claims 40-70 and 72-88 as being unpatentable over Park et al., in view of Mitchell et al., Harrison, Schultz and further in view of Gossen et al.

Conclusion

In view of the above remarks and the amendments to the claims, it is believed that this application is in condition for allowance. If a telephone conversation with Applicant's Attorney would expedite prosecution of the above-identified application, the Examiner is urged to call the undersigned at (617) 832-1000.

Respectfully submitted,
FOLEY, HOAG & ELIOT LLP

By: _____

One Post Office Square
Boston, MA 02109
Telephone: (617) 832-1000
Facsimile: (617) 832-7000

Matthew P. Vincent, Ph.D.
Registration No. 36,709
Attorney for Applicants

Date: February 4, 2000

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4892–4896, June 1993
Biochemistry

EXHIBIT A

Design superiority of palindromic DNA sites for site-specific recognition of proteins: Tests using protein stitchery

CHANGMOON PARK*†, JUDY L. CAMPBELL†‡, AND WILLIAM A. GODDARD III*†§

*Materials and Molecular Simulation Center, Beckman Institute (139-74), †Division of Chemistry and Chemical Engineering, and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by William A. Goddard III, February 12, 1993

ABSTRACT **Using protein stitchery with appropriate attachment of cysteines linking to either C or N termini of the basic region of the v-Jun leucine zipper gene-regulatory protein, we constructed three dimers—pCC, pCN, and pNN. All three bind specifically to the appropriately rearranged DNA recognition sites for v-Jun: ATGAcgTCAT, ATGAcgATGA, and TCATcgTCAT, respectively (K_d, ≈4 nM at 4°C). Results of DNase I footprinting provide strong support for bent recognition helices in leucine zipper protein–DNA complexes. Comparison of the results for pCC and pNN with those for pCN shows the design superiority of palindromic sequences for protein recognition.**

The mechanism by which cells respond to external stimuli is a fundamental problem in modern biology. Transcriptional regulatory proteins are known to play a key role in several systems evolved by cells to convert extracellular signals into altered gene expression (1). They operate by specifically binding to DNA target sites, which regulate the transcription of particular genes. Prominent among transcriptional regulatory proteins are the leucine zipper family of proteins, which recognize the DNA binding site as either homodimers or heterodimers (2–4).

The leucine zipper proteins are characterized by two functional segments: (*i*) the leucine zipper region, a helical region containing four or five leucines spaced at seven-amino acid intervals, and (*ii*) the basic region containing many basic residues (5–10). The basic region appears to be unfolded in solution but assumes an α-helical structure binding to its recognition site (11–13). Site-directed mutagenesis (6, 7) and domain swapping (8–10) experiments show that the leucine zipper region mediates dimerization and that the basic region is responsible for DNA binding. Experiments replacing the leucine zipper region with a three-peptide linker (14, 15) showed that the dimerized basic region recognizes the same site as the native protein, supporting the scissors grip model (5), where each monomer recognizes the half site of the symmetrical DNA binding site. Recently, we showed that the normal dimer (denoted pCC), which selectively recognizes the sequence ATGAcgTCAT, can be inverted to form a protein (denoted pNN) that selectively recognizes the inverted site, TCATcgATGA (15).

Gel electrophoresis experiments (22) with Jun homodimer and with Jun–Fos heterodimer showed that Jun and Fos induce DNA bending in the opposite direction upon binding to the specific site. To explain this, it was proposed that the basic region of jun has a bent α-helix, while the basic region of Fos has a straight helix. However, a recent x-ray crystal structure (21) for the complex between GCN4 and DNA containing the GRE site (ATGACTCAT) showed a straight single α-helix for the basic region of GCN4. Our current results support the interpretation that the v-Jun homodimer bound to its specific site has bent α-helices.

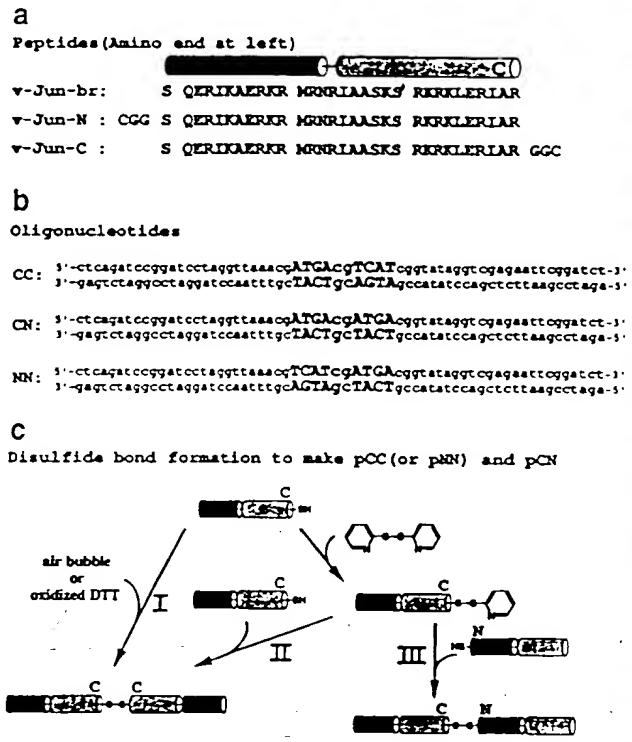

FIG. 1. Sequences of protein (*a*) and oligonucleotides (*b*) used in gel-retardation and footprinting studies. Total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. CGG is added to the N terminus of v-Jun-br to make v-Jun-N and GGC is added to the C terminus of v-Jun-br to make v-Jun-C. Proteins were chemically synthesized and checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology (15). (*c*) Strategy for making pCC (and pNN) and pCN dimers. v-Jun-C was incubated with 10 mM dithiothreitol (DTT) for 5 hr at room temperature and purified directly into 10 mM 2,2′-dithiodipyridine/100 mM sodium phosphate, pH 5.5, containing 30% acetonitrile. Resulting thiopyridyl-(v-Jun-C) was purified by HPLC. Purified monomer v-Jun-N underwent reaction with 2 equivalents of thiopyridyl-(v-Jun-C) in solution containing 100 mM tetraethylammonium acetate buffer (pH 7.5) and 15% acetonitrile for 12 hr at room temperature. The final product, pCN, was purified by HPLC (15).

Peptide Design

Using protein stitchery, we have made three kinds of v-Jun (16, 17) homodimers (denoted pCC, pNN, and pCN) and show here that each selectively recognizes the appropriately reorganized DNA binding sites ATGAcgTCAT, TCATcgATGA, and ATGAcgATGA (see Fig. 1). The concept of protein stitchery (15) is that the individual basic arms (half

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§To whom reprint requests should be addressed at •.

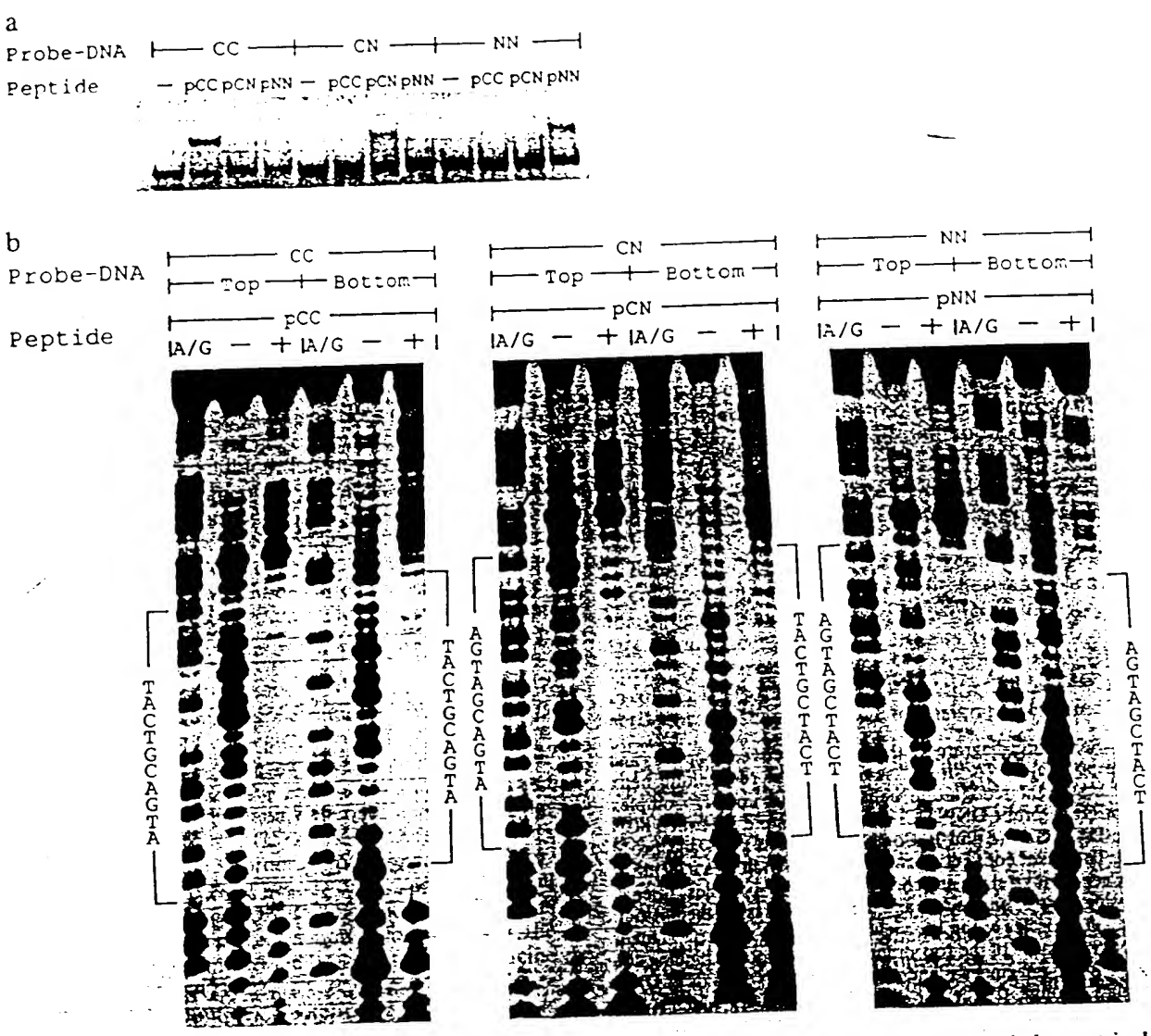

FIG. 2. (*a*) Gel-retardation assay for binding of pCC, pCN, and pNN to the CC, CN, and NN probe DNAs. Binding solution contains bovine serum albumin at 50 μg/ml, 10% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 4 mM $MgCl_2$, and the appropriate peptide at 3 nM in a 10-μl reaction volume. After 5000 cpm of each 5′ ^{32}P-labeled probe DNA was added, the solutions were stored at 4°C for 40 min and loaded directly on an 8% nondenaturing polyacrylamide gel in Tris/EDTA buffer at 4°C. As determined by titration of the gel shift, $K_d \approx 2$ nM for pCC/CC, $K_d \approx 6$ nM for pCN/CN, and $K_d \approx 4$ nM for pNN/NN, all at 4°C. These results show that each peptide binds specifically to its proposed binding site and not to the other sites. (*b*) DNase I footprinting assay of pCC, pCN, and pNN peptide with DNA containing the predicted binding sites for pCC, pCN, and pNN, respectively. Footprinting assay solution (in 50 μl) contains bovine serum albumin at 50 μg/ml, 5% (vol/vol) glycerol, 20 mM Tris·HCl (pH 7.5), 4 mM KCl, 2 mM $MgCl_2$, 1 mM $CaCl_2$, 20,000 cpm of each 5′ ^{32}P-labeled probe DNA, and the appropriate peptide at 50 nM. This solution was stored at 4°C for 40 min. After 5 μl of DNase I diluted in 1× footprinting assay buffer was added, the solutions were stored 1 min more at 4°C. DNase I digestion was stopped by addition of 100 μl of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, and sonicated salmon sperm DNA at 40 μg/ml. This mixture was phenol/chloroform-extracted, ethanol-precipitated, and washed with 70% (vol/vol) ethanol. The pellet was resuspended in 5 μl of formamide loading buffer, denatured at 90°C for 4 min, and analyzed on 10% polyacrylamide sequencing gel (50% urea). These results show that each peptide specifically binds to the proposed binding site and protects the whole site except for the case of pCN/CN, which shows some incomplete protection on the binding site. This exception is explained as due to binding to semispecific (half) sites by single arms as discussed in the text (see Fig. 4).

sites) of the dimer and the individual half sites of the DNA can be recombined or stitched together in various sequences to form new proteins selective for binding to the new DNA sites. Thus, we use here the recognition helix v-Jun-br of Fig. 1*a* with a cysteine linker at either the N (v-Jun-N) or the C (v-Jun-C) terminus. These can be combined to form either pNN, pCC, or pCN dimers as illustrated in Fig. 1*c*. Formation of pNN and pCC (via pathway I) is straightforward since each involves dimerization of identical monomers. To ensure formation of pCN, the cysteine at the C terminus of v-Jun-C was reacted with excess 2,2′-dithiodipyridine to form thiopyridyl-(v-Jun-C) (18, 19) and then coupled with the cysteine at the N terminus of v-Jun-N to form the pCN dimer (v-Jun-C)-S–S-(v-Jun-N) (pathway III; Fig. 1*c*). We also verified pathway II for forming pCC.

Results and Discussion

We carried out gel-retardation assays (15) for each of the three peptide dimers with oligonucleotides (Fig. 1*b*) corresponding to each of the three proposed binding sites. These results (Fig. 2*a*) show that each dimer recognizes the appropriate binding site specifically with no detectable binding to the other sites. It is important to note that this strong preference for dimer occurs even though all oligonucleotides contain proper sites for binding a single arm of each dimer.

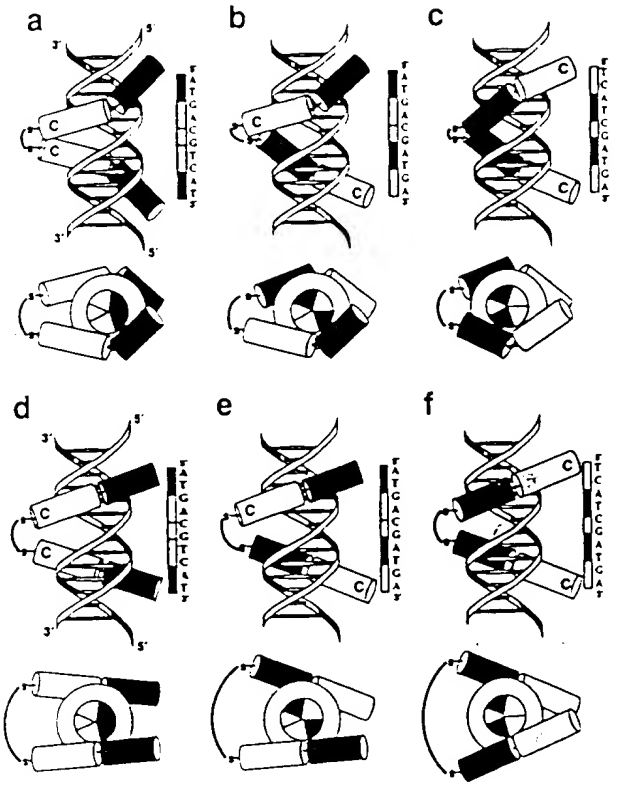

FIG. 3. Schematic diagram for the complex between peptides and their corresponding DNA sites assuming a bent recognition helix (*a–c*) and a straight recognition helix (*d–f*). (*a* and *d*) Complex between pCC and probe DNA CC. (*b* and *e*) Complex between pCN and probe DNA CN. (*c* and *f*) Complex between pNN and probe DNA NN. The linker connecting two monomers indicates a disulfide bond between the cysteines attached to the end of peptides. In each case, the side view is on the top and the top view is on the bottom. Outer and inner circles of the top view represent the outer and inner major groove surfaces of the top strand for the proposed binding site projected onto an imaginary plane perpendicular to the axis of DNA and running through the center of the peptide and binding site. Shading is used with the peptide and DNA contacts to ease the tracking of these regions in different cases. This diagram shows that a bent recognition helix can contact the same 4 bases for all three peptide dimers, while a linear recognition helix would contact different bases in the three peptide dimers. (This diagram is not meant to imply an exact correlation between where the basic region is bent and where the bases are positioned.)

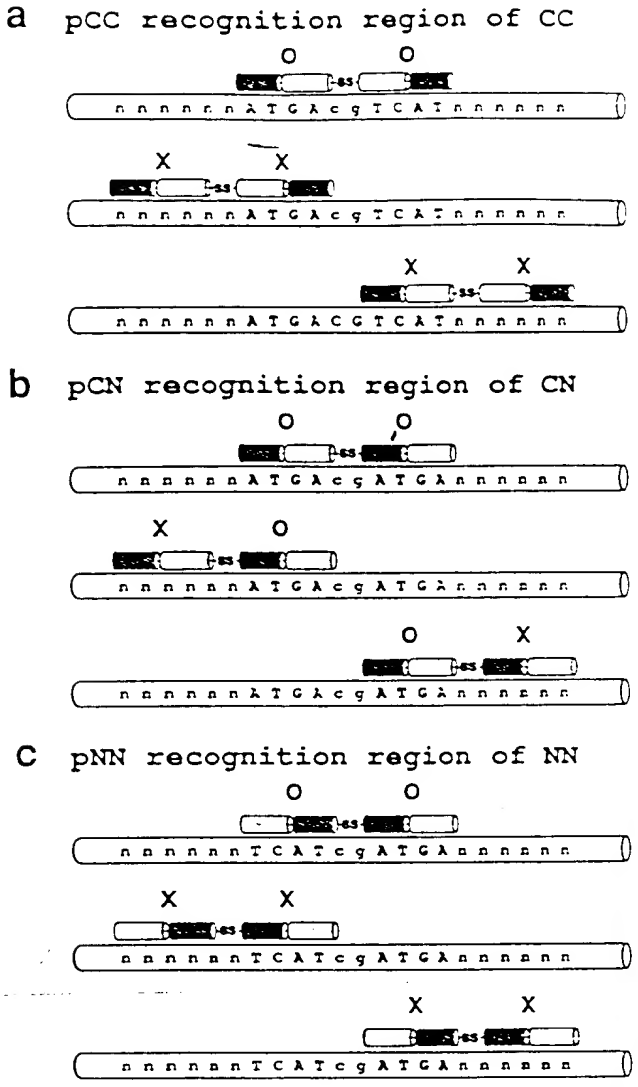

FIG. 4. Specific binding of protein at and near the corresponding DNA binding site. (*a*) Complex pCC/CC. (*b*) Complex pCN/CN. (*c*) Complex pNN/NN. O represents specific binding; X represents nonspecific binding. pCN/CN has one specific binding site and two sites for semispecific (half site) binding near its (nonpalindromic) binding site. However, pCC/CC and pNN/NN do not allow semispecific binding near their (palindromic) binding sites.

Therefore, at 3 nM peptide concentration the dimer does not make a stable complex with DNA unless both arms in the dimer recognize their proper sites. This implies cooperation between the monomers in recognizing the binding site (20). Since all three dimers have similar (2–6 nM) binding affinities with their own sites and since all three lead to the same length region protected from DNase I digestion (see below), we conclude that (*i*) all three cases involve similar conformations in the complex between DNA and peptide, and (*ii*) the monomer arm retains the same contact region in various dimers; this occurs despite the changing orientation of the monomers in the various peptide dimers (15).

There are two major models for the bound conformation of leucine zipper protein to the specific site. One is the induced helical fork model (13), which proposes a straight single α-helix for the basic region, and the other is a scissors grip model (5) which proposes a bent α-helix for the basic region. The recent x-ray crystal structure (21) for the complex of GCN4 containing only the basic and leucine zipper region and DNA-containing GRE site showed that the basic region of each protein has a straight α-helix conformation recognizing each half site of the dimer binding site. There was no DNA bending caused by protein binding (21). However, there remain many problems with assuming that the basic region is in all cases a straight α-helix: (*i*) The bases flanking the active site affect the binding of leucine zipper protein even though the crystal structure shows no direct contacts with protein (21). (*ii*) Gel electrophoresis experiments using Jun homodimer and Jun-Fos heterodimer showed that Jun and Fos induce DNA bending in opposite directions upon binding to their site (22), whereas GCN4 does not induce DNA bending (21, 27). (*iii*) Even though GCN4-br (a peptide containing the basic region of GCN4 protein) showed no specific binding (for details see ref. 14), we find that the monomer v-Jun-br (a peptide containing only the basic region of v-Jun; see Fig. 1*a*) specifically binds to the dimer site and shows the same protection as the dimer. Our conclusion then is that there is no universal model for the DNA-bound conformation of the basic region of leucine zipper proteins. Whether it is linear (as in GCN4) or bent (as in Jun) depends on the specific primary sequence and the properties of the solutions (stabilizers, pH, etc.) used in the experiments.

The result that all three dimers (pCC, pNN, and pCN) bind strongly to the appropriate combination of oligonucleotide sites implies that the helical binding arm is bent (5, 22) (see Fig.). Our argument is as follows. The optimum binding site for both Jun homodimer and the Jun–Fos heterodimer is known to e ATGAcTCAT or ATGAcgTCAT, where the inner 7 or 8 ases play an important role in recognition (23, 24). The x-ray rystal structure of GCN4 bound to DNA leads to straight -helices, which have direct contacts with only the inner 7 ases of the GRE site (ATGACTCAT). Thus, each arm ecognizes the half-site (gATGAc or gTCATc) of the dimer inding site asymmetrically. If the same were true for v-Jun nd if the same contacts are maintained between the protein nd bases for the bound conformations of pCC/CC, pNN/NN, nd pCN/CN (as expected since the binding constants and protection are the same), then the orientations of the binding arms would have very different orientations (Fig. 3 *e* and *f*). This should result in different protection from DNase I digestion (not observed). In addition, for the pNN/NN complex, this would lead to N termini of the two arms too distant to be connected by the added linker, GGCCGG. The alternative to Fig. 3 *d–f* is for each dimer to have the same angle (as in Fig. 3*d*). Thus, the actual contact region would not be equivalent in the three cases and it would be difficult to explain the gel retardation and footprinting results. Thus, we conclude that for v-Jun the basic region becomes bent upon binding to the DNA.

On the other hand, with the recognition helix bent roughly at the middle of the helix (as indicated in Fig. 3 *a–c*), it is plausible that the contact regions are ATGAcgTCAT for pCC, TCATcgATGA for pNN, and ATGAcgATGA for pCN. This leads to equivalent contact regions in all three cases and to the roughly equivalent binding energies apparent in Fig. 2*a*. In addition, footprinting (15) of the three peptide dimers (Fig. 2*b*), each with the appropriate oligonucleotide dimer, suggests that the complexed peptide dimers protect the full specific site (all 10 bp) from DNase I digestion. These results strongly support the bent recognition helix model for the basic dimers considered here and hence also for the leucine zipper parent proteins (21, 22).

For the pCN/CN complex, footprinting (Fig. 2*b*) shows incomplete protection on the binding site and partial protection on the bases flanking the binding site, whereas for pCC/CC and pNN/NN this does not happen. This occurs even though gel-retardation assays indicate specific binding for all complexes. Our explanation of this (Fig. 4) suggests why palindromic sequences are so common for selective binding of regulatory proteins (25, 26). This reasoning is supported by recent results we have observed showing that (*i*) the monomer of v-Jun containing only the basic region (v-Jun-br) specifically protects both pCC and pNN binding sites identically to the protection provided by the dimers pCC and pNN, respectively; (*ii*) at 3 nM concentration, gel retardation showed that pCC (and pCN) has lower binding affinity for the DNA probe carrying a sequence of cgATGAcgTCATcgTCATcg (containing pCC and pCN binding sites overlapping half of each dimer binding site in the center) than for CC (and CN) probe DNA. These results imply that the half site, gTCATc (or gATGAc), added next to the pCC (or pCN) binding site interferes with the binding of pCC (or pCN) to the dimer binding site (because the half site can be used as a binding site for each arm of the dimer if the orientation between the site and arm fits). Details of these results will be published elsewhere. Fig. 4 indicates the strength of binding for all three peptide dimers at or near their DNA recognition sites. Here, O represents good binding, while X represents nonspecific binding. The palindromic sites for pNN and pCC lead to binding only when the protein is exactly at the recognition site, whereas pCN can recognize both full site (both arms bound) and half sites (one arm bound). In gel retardation and DNase I footprinting, semispecific binding competes with specific binding. This occurs because one arm of the semispecifically bound peptide would cover half of the specific binding site, preventing another dimer from binding and providing full protection. This explains (*i*) why gel-retardation assays (Fig. 2*a*) show lower binding affinity for the pCN/CN complex compared to the pCC/CC and pNN/NN complexes and (*ii*) why footprinting assays (Fig. 2*b*) show incomplete protection on the binding site and partial protection on a few bases flanking the binding site. Such semispecific binding interferes with the site-specific binding and would eventually result in low production and abnormally slow growth. However, gel retardation shows no detectable nonspecific or semispecific binding at low peptide concentration, indicating that semispecific binding is significantly weaker than specific binding. After dimerization, the proteins suitable for palindromic dimer binding sites avoid semispecific DNA binding, leading to more selective recognition of the specific sites. Thus, palindromic dimer binding sites provide a good design for selective molecular recognition and for further flexibility the link can align sites (Fig. 3) to modify recognition.

The results on the three dimers considered here provide encouragement that this protein stitchery approach is feasible for designing and synthesizing proteins to recognize long DNA sequences. Thus, for trimers to recognize 15-bp sequences, we are using an approach similar to that of Fig. 1*c* involving appropriate use of cysteine linkages and transfer activators. It seems possible to design proteins for 20 bp and longer.

In summary, we find the following: (*i*) Protein stitchery of v-Jun leads to three dimers (pCC, pNN, and pCN), each of which binds specifically to the appropriate rearrangement of DNA sites. Thus, there is cooperation between the two monomers of the dimer in binding to DNA, which depends on the relative orientation of two monomers in the dimer. (*ii*) These results provide strong support for the bent α-helix model of the basic region when bound to DNA. (*iii*) These results provide an explanation for the advantage of dimerization and the use of palindromic sites in the site-selective binding of proteins to DNA. (*iv*) These results show protein stitchery to be useful for establishing the conformation and mechanism for binding of proteins to their DNA binding sites.

This is contribution no. 8776 from the Division of Chemistry and Chemical Engineering, California Institute of Technology. We thank Prof. Peter Kim for helpful criticisms. This work was supported by a grant from the Department of Energy-Advanced Industrial Concepts Division. The facilities of the Materials and Molecular Simulation Center are also supported by grants from the National Science Foundation-CHE 91-100289), the National Science Foundation-Advanced Computing, the Department of Energy-Chemical Biology, Allied Chemical, Asahi Chemical, Asahi Glass, Chevron, BF Goodrich, BP America, and Xerox.

1. Curran, T. & Franza, B. R., Jr. (1988) *Cell* 55, 395–397.
2. Kouzarides, T. & Ziff, E. (1989) *Nature (London)* 340, 568–571.
3. Turner R. & Tjian, R. (1989) *Science* 243, 1688–1694.
4. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* 243, 1695–1699.
5. Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* 246, 911–916.
6. Neuberg, M., Schermann, M., Hunter, J. B. & Muller, R. (1989) *Nature (London)* 338, 589–590.
7. Ransone, L. J., Visvader, J., Wamsley, P. & Verma, I. M. (1990) *Proc. Natl. Acad. Sci. USA* 87, 3806–3810.
8. Ransone, L. J., Wamsley, P., Mosley, K. L. & Verma, I. M. (1990) *Mol. Cell. Biol.* 10, 4565–4573.
9. Agre, P., Johnson, P. K. & McKnight, S. L. (1989) *Science* 246, 922–926.
10. Neuberg, M., Adamkiewicz, J., Hunter, J. P. & Muller, R. (1989) *Nature (London)* 341, 243–245.
11. Weiss, M. A. (1990) *Biochemistry* 29, 8020–8024.

12. Patel, L., Abate, C. & Curran, T. (1990) *Nature (London)* **347**, 572–575.
13. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249**, 774–778.
14. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249**, 769–771.
15. Park, C., Campbell, J. L. & Goddard, W. A., III (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9094–9096.
16. Kaki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2848–2852.
17. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P. K. (1988) *Cell* **52**, 705–712.
18. Corey, D. & Schultz, P. G. (1987) *Science* **238**, 1401–1403.
19. Zuckermann, R., Corey, D. & Schultz, P. G. (1987) *Nucleic Acids Res.* **15**, 5305–5321.
20. Abate, C., Luk, D., Gagne, E., Roeder, R. G. & Curran, T. (1990) *Mol. Cell. Biol.* **10**, 5532–5535.
21. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) *Cell* **71**, 1223–1237.
22. Kerpolla, T. K. & Curran, T. (1991) *Science* **254**, 1210–1214.
23. Ryseck, R. & Bravo, R. (1991) *Oncogene* **6**, 533–542.
24. Schule, R., Umesono, K., Mangelsdorf, D. J., Bolado, J., Pike, J. W. & Evans, R. M. (1990) *Cell* **61**, 497–504.
25. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
26. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
27. Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6034–6038.

EXHIBIT B

2ND STORY of Level 1 printed in FULL format.

Copyright 1995 Newsday, Inc.
Newsday

January 6, 1995, Friday, NASSAU AND SUFFOLK EDITION

SECTION: NEWS; Pg. A17

LENGTH: 260 words

HEADLINE: Scientists Flip on Gene 'Switch'

BYLINE: By Robert Cooke. STAFF WRITER

BODY:

Custom-made molecular switches that can turn genes on or off on cue are now being created in the laboratory, scientists said yesterday, showing promise as a powerful new way to attack genetic diseases, cancer and perhaps even AIDS.

The new approach is analogous to making an artificial key to start a stalled car. By finding a way to turn on a dormant gene that makes normal hemoglobin, for example, it may become possible to rescue people with sickle cell anemia and similar blood disorders.

The same technique may also be used to block gene activity, stopping the wildfire growth of cancer cells, or to shut off a virus' genes to abort an infection. An obvious target would be HIV, which causes AIDS.

According to biologist Carl Pabo, at the Massachusetts Institute of Technology, laboratory tests have proved the artificial switch can find and control a single gene among the 80,000 that exist in humans. The protein molecule they created attached itself at exactly the right place among 3 billion links in the long, complex chain of human DNA.

"The critical thing was showing it can find the proper site," Pabo said. But, he cautioned, much work must yet be done before any therapy for human diseases becomes possible.

In the tests, the switch worked as hoped, activating the desired gene. It represents a "fundamental advance and a very powerful method," Pabo said.

A full report on the experiment was published in Science yesterday. Other members of the team were Joel Pomerantz, a graduate student, and biologist Philip Sharp, both at MIT.

LANGUAGE: ENGLISH

LOAD-DATE: January 07, 1995

10. At 240°C, the highest temperature investigated, the pressure generated in these tubes amounted to only 33 atm, too low to produce a significant effect on the rates of known reactions in solution [for a survey, see W. J. le Noble, *Prog. Phys. Org. Chem.* **5**, 207 (1967)]. The sealed quartz tubes (1 cm in outside diameter, with a 2-mm wall) used in these experiments proved superior to sealed borosilicate tubes in strength and resistance to solvent attack. In addition, quartz tubes allowed the progress of decarboxylation of 1-methylorotate to be monitored without the opening of the tubes by observation of the ultraviolet absorption spectrum of the contents at intervals by means of a diode array spectrophotometer.
11. The absence of catalysis by methylamine is of special interest, in view of the fact that lysine is the single active site residue whose presence has been identified as essential for the catalytic activity of OMP decarboxylase [J. A. Smiley and M. E. Jones, *Biochemistry* **31**, 12162 (1992)].
12. This rate constant is smaller than a value extrapolated earlier (5 × 10[illegible] s[illegible]) for 1,3-dimethylorotic acid at 37°C from its behavior in sulfolane at 20[illegible] and 226°C (6). This difference in rates is probably due at least in part to differences in substrate structure.
13. K. J. Laidler, *Chemical Kinetics* (McGraw-Hill, New York, ed. 2, 1965), pp. 498–53[illegible].
14. Linear Arrhenius plots are almost invariably observed for simple chemical reactions (13), but departures from linearity have been reported occasionally for more complex physiological processes in which the mechanism or the rate-determining step changes with temperature [for a review, see F. H. Johnson, H. Eyring, M. J. Polissar, *The Kinetic Basis of Molecular Biology* (Wiley, New York, 1954), pp. 187–[illegible]]. In those few cases in which nonlinearity has been observed, the slope at lower temperature is greater than the slope at higher temperature. If the same were true of the simple reaction described here (an unlikely possibility that cannot be tested at present), then the rate of reaction at 25°C would have been overestimated by the present extrapolations.
15. R. W. Tucker, E. E. Hazen Jr., F. A. Cotton, *Mol. Cell. Biochem.* **22**, 67 (1978).
16. C. A. Bunton, M. M. Mhala, K. G. Oldham, C. A. Vernon, *J. Chem. Soc.* **1960**, 3293 (1960).
17. At temperatures above 100°C, hydrolysis of dimethyl phosphate proceeds even more slowly than does decarboxylation of 1-methylorotic acid (Fig. [illegible]). The entropy of activation for decarboxylation is positive (+19.7 cal deg[illegible] mol[illegible]), whereas that of phosphodiester hydrolysis is negative (−27.6 cal deg[illegible] mol[illegible]), so that decarboxylation proceeds much more slowly than does phosphodiester hydrolysis in neutral solution at room temperature. Values of these entropies of activation are in the ranges typically observed for mono- and bimolecular reactions, respectively (13).
18. E. H. Serpersu, D. Shortle, A. S. Mildvan, *Biochemistry* **25**, 68 (1986).
19. If there is a difference in mechanism between the enzymatic reaction and the nonenzymatic reaction or if the rate of the enzyme reaction is limited by some event that does not involve bond making or breaking in substrates, such as product release, then these values represent upper limits of the dissociation of the enzyme substrate complex in the transition state [for a review, see R. Wolfenden, *Annu. Rev. Biophys. Bioeng.* **5**, 271 (1976)].
20. Table 1 is restricted to enzyme reactions that do not appear to proceed by double displacement mechanisms [D. E. Koshland Jr., *Biol. Rev. Camb. Philos. Soc.* **28**, 416 (1953)], because comparisons of enzymatic and nonenzymatic reaction rates do not permit straightforward estimation of transition state affinity for reactions of this type (19). Large rate enhancements, although not as large as that for OMP decarboxylase, have also been reported for certain multisubstrate reactions, including those catalyzed by catechol O-methyltransferase [[illegible], J. K. Coward, R. L. Schowen, *J. Am. Chem. Soc.* **101**, 4342 (1979)] and [illegible] [J. W. [illegible] and W. J. R[illegible], *Biochemistry* **23**, [illegible] (1984)].
21. J. A. A[illegible], J. B. Bell, M. E. Jones, R. W. [illegible], *Biochemistry* **29**, [illegible].
22. J. A. Smiley, P. Paneth, M. H. O'Leary, J. B. Bell, M. E. Jones, *ibid.* **30**, 6216 (1991).
23. [illegible] inhibitor with an inhibition constant value of [illegible] M is bound noncovalently [21] 4 to 5 orders of magnitude more tightly than is the substrate or product, probably because of its resemblance to a nitrogen ylide intermediate in OMP decarboxylation [H. L. Levine, R. S. Brody, F. H. Westheimer, *ibid.* **19**, 4993 (1980)].
24. Auxotrophs of *Escherichia coli* have been used to prepare an antibody catalyzing the decarboxylation of orotic acid [J. A. Smiley and S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8319 (1994)]. The present results indicate that this antibody achieves a remarkable catalytic proficiency of approximately 10^{15} M[illegible], although yeast OMP decarboxylase exceeds that value by a factor of roughly [illegible].
25. E. R. Garrett and P. J. Mehta, *J. Am. Chem. Soc.* **94**, 8532 (1972).
26. D. P. Dumas, S. R. Caldwell, J. R. Wild, F. M. Raushel, *J. Biol. Chem.* **264**, 19659 (1989).
27. [illegible] *Biochemistry* **30**, [illegible].
28. [illegible] **14**, [illegible].
29. E. H. Serpersu, D. Shortle, A. S. Mildvan, *ibid.* **25**, [illegible] (1986).
30. S. W. Hunt and P. A. Hoffee, *J. Biol. Chem.* **257**, 14240 (1982).
31. W. E. DeWolf Jr., F. A. Fullin, V. L. Schramm, *ibid.* **254**, 10868 (1979).
32. J. R. Whitaker, G. Menger, M. L. Bender, *Biochemistry* **5**, 386 (1966).
33. R. M. Pollack, B. Zeng, J. P. G. Mack, S. Eldin, *J. Am. Chem. Soc.* **111**, 6419 (1989).
34. S. J. Putman, *Biochem. J.* **129**, 301 (1972).
35. H. Steiner, B.-H. Jonsson, S. Lindskog, *Eur. J. Biochem.* **59**, 253 (1975).
36. J. L. Kofron, P. Kuzmic, V. Kishore, E. Colon-Bonilla, D. H. Rich, *Biochemistry* **30**, 6127 (1991).

37. Supported by NIH grant GM-18325. We are grateful to S. Rowley for technical assistance.

22 June 1994; accepted 29 September 1994

Structure-Based Design of Transcription Factors

Joel L. Pomerantz, Phillip A. Sharp, Carl O. Pabo

Computer modeling suggested that transcription factors with novel sequence specificities could be designed by combining known DNA binding domains. This structure-based strategy was tested by construction of a fusion protein, ZFHD1, that contained zinc fingers 1 and 2 from Zif268, a short polypeptide linker, and the homeodomain from Oct-1. The fusion protein bound optimally to a sequence containing adjacent homeodomain (TAATTA) and zinc finger (NGGGNG) subsites. When fused to an activation domain, ZFHD1 regulated promoter activity in vivo in a sequence-specific manner. Analysis of known protein-DNA complexes suggests that many other DNA binding proteins could be designed in a similar fashion.

Transcription factors are critical regulators of gene expression. The rational design of transcription factors with novel DNA binding specificities and regulatory activities will provide reagents for both biological research and gene therapy. The recent determination of a series of structures of protein-DNA complexes has facilitated a design strategy that uses computer modeling to predict how DNA binding domains could be combined to generate novel specificities. We explored this strategy by designing and testing a zinc finger–homeodomain fusion protein.

Computer modeling studies were used to visualize how zinc fingers might be fused to the Oct-1 homeodomain. The known crystal structures of the Zif268-DNA (1) and Oct-1–DNA (2) complexes were aligned by superimposition of the double helices in several different registers. Two arrangements were particularly interesting. In one alignment, the COOH-terminal end of zinc finger 2 was 8.8 Å away from the NH_2-terminal arm of the homeodomain (Fig. 1), which suggested that a short polypeptide linker could connect these domains. In this model, the fusion protein would bind a hybrid DNA site with the sequence 5′-AAATNNTGGGCG-3′. The Oct-1 homeodomain would recognize the AAAT subsite, zinc finger 2 would recognize the TGG subsite, and zinc finger 1 would recognize the GCG subsite. There was no possibility for steric interference between the zinc fingers and the homeodomain in this arrangement. Superimposition of the DNA duplexes in other registers generated a second plausible arrangement for a hybrid protein (3); however, this model was not as favorable because there was a risk of steric interference between the zinc fingers and the homeodomain.

The design strategy was tested by construction of a fusion protein, ZFHD1, that contained fingers 1 and 2 of Zif268, a gly-gly-arg-arg linker, and the Oct-1 homeodomain (Fig. 2A). A glutathione-S-transferase (GST) domain was added to facilitate expression and purification, and the DNA binding activity of this fusion protein was

J. L. Pomerantz, Center for Cancer Research, Department of Biology, and Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
P. A. Sharp, Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
C. O. Pabo, Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA [illegible].

determined by selection of binding sites from a random pool of oligonucleotides. After four rounds of selection, 16 sites were cloned and sequenced (Fig. 2B). Comparison of these sequences revealed the consensus binding site 5'-TAATTANGGGNG-3' (Fig. 2C). The 5' half of this consensus, TAATTA, resembled a canonical homeodomain binding site (TAATNN) (4) and matched the site (TAATNA) that is preferred by the Oct-1 homeodomain in the absence of the POU-specific domain (5). The 3' half of the consensus, NGGGNG, resembled adjacent binding sites for fingers 2 (TGG) and 1 (GCG) of Zif268. The guanine residues were more tightly conserved than were the other positions in these zinc finger subsites, and the crystal structure shows that these are the positions of the critical side-chain–base interactions (1).

The ZFHD1 consensus sequence (5'-TAATTANGGGNG-3') matched the model that appeared to be most structurally feasible (6), but because of the internal symmetry of the TAATTA subsite, this sequence was also consistent with the homeodomain binding in another orientation (Fig. 2D; compare mode 1 and mode 2). This alternative arrangement, in which the critical TAAT is on the other strand and is directly juxtaposed with the zinc finger (TGGGCG) subsites, was considered unlikely because modeling had suggested that this arrangement required a linker to span >20 Å between the COOH-terminus of finger 2 and the NH_2-terminus of the homeodomain. To determine how the homeodomain bound to the TAATTA sequence in the 5' half of the consensus, ZFHD1 was tested for binding to probes (5'-TAATGATGGGCG-3' and 5'-TCATTATGGGCG-3') designed to distinguish between these orientations. ZFHD1 bound to the 5'-TAATGATGGGCG-3' probe with a dissociation constant of 8.4×10^{-11} M and preferred this probe to the 5'-TCATTATGGGCG-3' probe by a factor of 33 (Fig. 3A; compare lanes 6 to 10 and 11 to

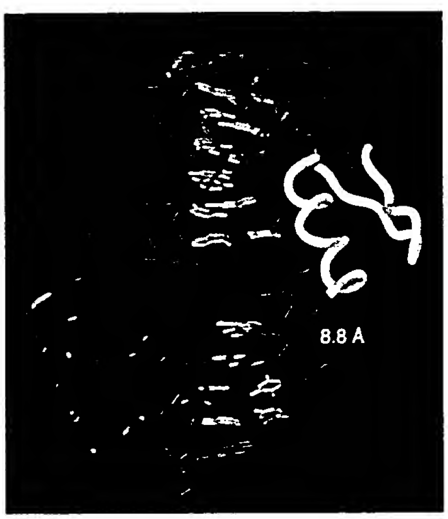

Fig. 1. Model of a zinc finger–homeodomain hybrid. Finger 1 of Zif268 is depicted in purple, finger 2 in yellow, and the Oct-1 homeodomain in red. The DNA is blue, with the base pairs in the AAAT and TGGGCG subsites highlighted in cyan; the hybrid protein recognizes a sequence of the form 5'-AAATNNTGGGCG-3'. The C_α of Gly (COOH terminus of finger 2) is 8.8 Å from the C_α of Arg [illegible] in the crystal structure [illegible]. The figure was generated with [illegible] San Diego [illegible].

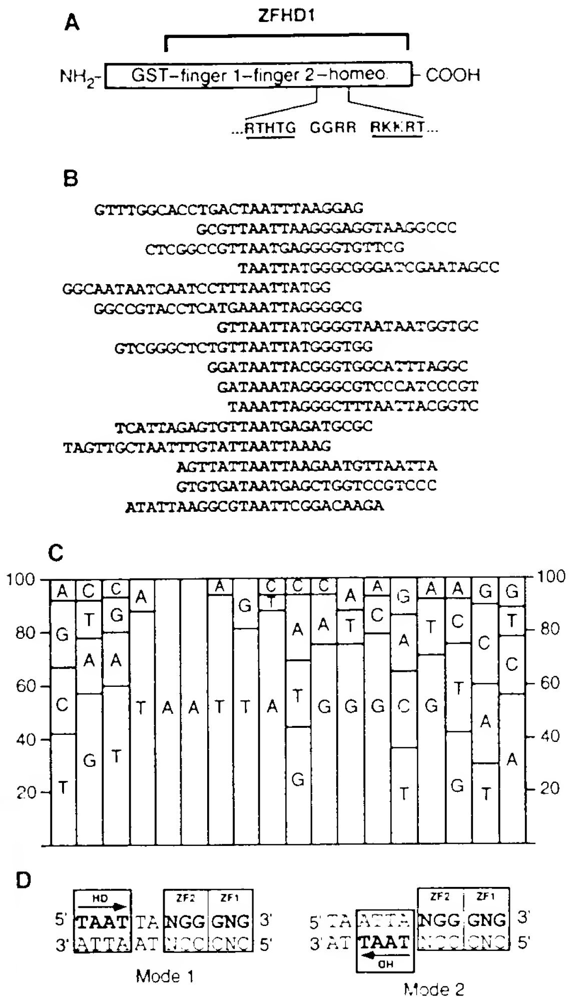

Fig. 2. Selection by ZFHD1 of a hybrid binding site from a pool of random oligonucleotides. (**A**) Structure of the fusion protein used to select binding sites (19). The underlined residues are from the Zif268-DNA (1) and Oct-1-DNA (2) crystal structures and correspond to the termini used in the computer modeling studies. The linker contains two glycine residues that were included for flexibility and the two arginine residues that are present at positions −1 and 1 of the Oct-1 homeodomain. (**B**) Sequences of 16 sites isolated after four rounds of binding site selection (20). (**C**) Consensogram derived from the sequences in (B) that indicates the percent occurrence of each nucleotide at each position. (**D**) Schematic diagram illustrating the two possible orientations of the homeodomain subsite relative to the zinc finger subsite suggested by the consensus. Mode 1 corresponds to the arrangement shown in Fig. 1.

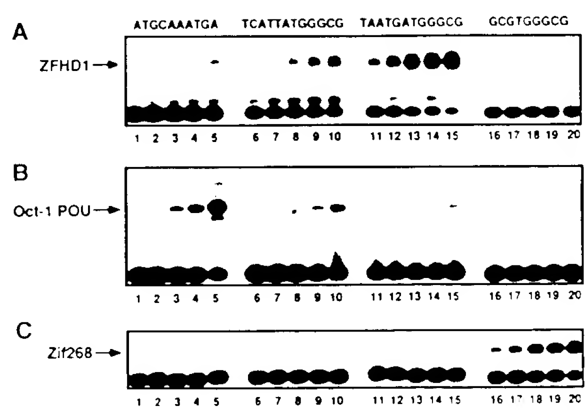

Fig. 3. Comparison of the DNA binding specificity of ZFHD1, the Oct-1 POU domain (which contains a homeodomain and a POU-specific domain), and the three zinc fingers from Zif268 (21). (**A**) The GST-ZFHD1 protein was titrated into DNA binding reactions containing the probe listed at the top of each set of lanes. Lanes 1, 6, 11, and 16 contain the protein at 9.8×10^{-11} M; protein concentration was increased in threefold increments in subsequent lanes of each set. The position of the protein-DNA complex is indicated by an arrow. (**B**) The PA–Oct-1 POU fusion protein (16) was titrated into parallel DNA binding reactions as in (A), but lanes 1, 6, 11, and 16 contain the protein at $2.1 \times 10^{-[illegible]}$ M. The position of the protein-DNA complex is indicated by an arrow. (**C**) A peptide containing Zif268 fingers 1, 2, and 3 [illegible] was titrated into parallel DNA binding reactions with lanes 1, 6, 11, and 16 containing the peptide at [illegible] M. The position of the protein-DNA complex is indicated by an arrow.

15). This suggested that the first four bases of the consensus sequence form the critical TAAT subsite that is recognized by the homeodomain (mode 1) and that ZFHD1 bound as predicted in the model (Fig. 1).

We compared ZFHD1, Oct-1, and Zif268 for their abilities to distinguish among the Oct-1 site 5′-ATGCAAATGA-3′, the Zif268 site 5′-GCGTGGGCG-3′, and the hybrid binding site 5′-TAATGATGGGCG-3′. The fusion protein ZFHD1 preferred the optimal hybrid site to the octamer site by a factor of 240 (Fig. 3A; compare lanes 1 to 5 and 11 to 15) and did not bind to the Zif site (lanes 16 to 20). The POU domain of Oct-1 (Fig. 3B) bound to the octamer site with a dissociation constant of 1.8×10^{-10} M (lanes 1 to 5), preferring this site to the hybrid sequences by factors of 10 (lanes 6 to 10) and 30 (lanes 11 to 15) and did not bind to the Zif site (lanes 16 to 20) (7). The three fingers of Zif268 (Fig. 3C) bound to the Zif site with a dissociation constant of 3.3×10^{-10} M (lanes 16 to 20) and did not bind to the other three sites (lanes 1 to 15). Thus ZFHD1 bound tightly and specifically to the hybrid site and displayed DNA binding specificity that was clearly distinct from that of either of the original proteins.

We fused ZFHD1 to a transcriptional activation domain and determined in transfection experiments whether the DNA binding protein could function in vivo. An expression plasmid encoding ZFHD1 fused to the COOH-terminal 81 amino acids of the herpes simplex virus VP16 protein (ZFHD1-VP16) was cotransfected into 293 cells with reporter constructs containing the SV40 promoter and the firefly luciferase gene. To determine whether the fusion protein could specifically regulate gene expression, we tested reporter constructs containing two tandem copies of the ZFHD1 site 5′-TAATGATGGGCG-3′, the octamer site 5′-ATGCAAATGA-3′, or the Zif site 5′-GCGTGGGCG-3′ inserted upstream of the SV40 promoter. When the reporter contained two copies of the ZFHD1 site, the ZFHD1-VP16 protein stimulated the activity of the promoter in a dose-dependent manner (8). Furthermore, the stimulatory activity was specific for the promoter containing the ZFHD1 binding sites (Fig. 4). At concentrations of protein that stimulated this promoter 44-fold, no stimulation above background was observed for promoters containing either the octamer or Zif sites. Thus, ZFHD1 efficiently and specifically recognized its target site in vivo.

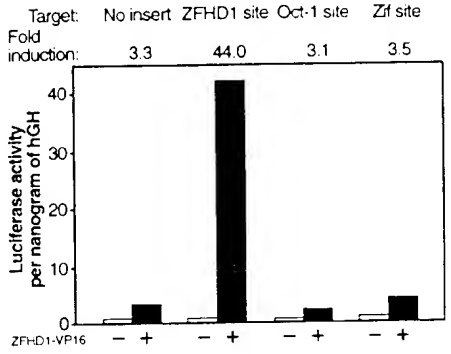

Fig. 4. Regulation of promoter activity in vivo by ZFHD1. The 293 cells were cotransfected with 5 μg of reporter vector, 10 μg of expression vector, and 5 μg of pCMV-hGH (gift of J. Parvin) used as an internal control (22). The reporter vectors contained two tandem copies of the ZFHD1 site (TAATGATGGGCG), the Oct-1 site (ATGCAAATGA), the Zif site (GCGTGGGCG), or no insert. The expression vector encoded the ZFHD1 protein fused to the COOH-terminal 81 amino acids of VP16 (plus signs), and the empty expression vector Rc/CMV was used as control (minus signs). The amount of luciferase activity obtained, normalized to hGH production, was set to 1.0 for the cotransfection of Rc/CMV with the no-insert reporter pGL2-Promoter. Bar graphs represent the average of three independent experiments. Actual values and standard deviation, reading from left to right, are: 1.00 ± 0.05, 3.30 ± 0.63; 0.96 ± 0.08, 42.2 ± 5.1; 0.76 ± 0.07, 2.36 ± 0.34; and 1.22 ± 0.10, 4.22 ± 1.41. Fold induction refers to the amount of normalized activity obtained with the ZFHD1-VP16 expression construct divided by that obtained with Rc/CMV.

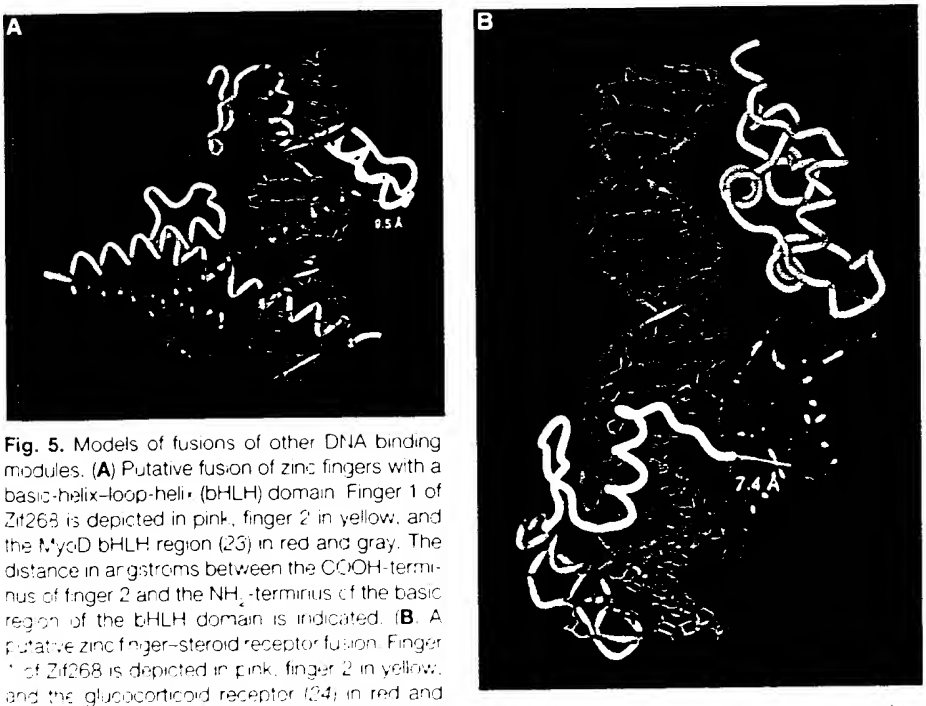

Fig. 5. Models of fusions of other DNA binding modules. (**A**) Putative fusion of zinc fingers with a basic-helix–loop–helix (bHLH) domain. Finger 1 of Zif268 is depicted in pink, finger 2 in yellow, and the MyoD bHLH region (23) in red and gray. The distance in angstroms between the COOH-terminus of finger 2 and the NH_2-terminus of the basic region of the bHLH domain is indicated. (**B**) A putative zinc finger–steroid receptor fusion. Finger 1 of Zif268 is depicted in pink, finger 2 in yellow, and the glucocorticoid receptor (24) in red and gray. The distance in angstroms between the COOH-terminus of finger 2 and the NH_2-terminus of the glucocorticoid receptor is indicated. These figures were generated with Insight II.

This structure-based strategy of fusing known DNA binding modules may provide a general method for designing transcription factors with novel DNA binding specificities. Computer modeling suggests a number of other plausible arrangements for hybrid proteins. Figure 5 illustrates models of a zinc finger–basic-helix–loop-helix fusion protein (Fig. 5A) and a zinc finger–steroid receptor fusion protein (Fig. 5B) that should recognize hybrid binding sites. In each case, the modules can be fused by a short polypeptide linker without steric interference between the domains. This strategy could also be extended by variation in the length and sequence of the polypeptide linkers and then use of selection methods to optimize the binding affinity and specificity of the hybrid protein.

The strategy of fusing modules can also be combined with strategies for changing the sequence specificity of individual modules. Several DNA binding domains are amenable to mutational strategies for changing sequence specificity (9, 10), and zinc fingers may offer the most versatility (11). Combining structure-based design with mutational changes in specificity would greatly expand the range of sequences that could be targeted by hybrid domains.

The high affinity of ZFHD1 for its optimal site and the fact that ZFHD1, Oct-1, and Zif268 all clearly preferred different sites illustrate the success of the combinatorial approach. The specificity of the hybrid transcription factor depends on the relatively moderate affinity, but high sequence specificity, for the binding of a single module and on the chelate effect (12)

provided by the covalent linkage of modules. The design criteria that allowed the construction of ZFHD1 included the short length of polypeptide linker that was required to fuse the DNA binding domains and the absence of steric interference between these domains.

Designed transcription factors will be useful for the targeted regulation of specific cellular genes. The use of particular DNA binding domains in a hybrid (or the addition of other domains) may allow a protein to interact with other cellular factors or to be modulated by a particular regulatory pathway. The structure-based design of hybrid transcription factors should facilitate the development of efficient and specific reagents for biological research and gene therapy.

REFERENCES AND NOTES

1. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
2. J. D. Klemm, M. A. Rould, R. Aurora, W. Herr, C. O. Pabo, *Cell* **77**, 21 (1994).
3. This alternative arrangement would also have a short (<10 Å) linker connecting zinc finger 2 to the homeodomain, but the subsites are arranged so that the predicted binding sequence is 5'-CGCCCATNAAAT-3'.
4. A. Laughon, *Biochemistry* **30**, 11357 (1991).
5. C. P. Verrijzer et al., *EMBO J.* **11**, 4993 (1992).
6. No selected sites matched the binding sequence predicted for the alternative arrangement (3).
7. The relatively high affinity of the POU domain for the hybrid sites may result from nonspecific contacts made by the POU-specific domain.
8. J. L. Pomerantz, unpublished results.
9. P. Youderian, A. Vershon, S. Bouvier, R. T. Sauer, M. M. Susskind, *Cell* **35**, 777 (1983); R. P. Wharton and M. Ptashne, *Nature* **316**, 601 (1985); M. Suckow, A. Madan, B. Kisters-Woike, B. von Wilcken-Bergmann, B. Müller-Hill, *Nucleic Acids Res.* **22**, 2198 (1994).
10. J. L. Pomerantz and P. A. Sharp, *Biochemistry* **33**, 10851 (1994).
11. E. J. Rebar and C. O. Pabo, *Science* **263**, 671 (1994); A. C. Jamieson, S.-H. Kim, J. A. Wells, *Biochemistry* **33**, 5689 (1994); R. Desjarlais and J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2256 (1993).
12. W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4046 (1981).
13. B. A. Christy, L. F. Lau, D. Nathans, *ibid.* **85**, 7857 (1988).
14. R. A. Sturm, G. Das, W. Herr, *Genes Dev.* **2**, 1582 (1988).
15. F. M. Ausubel et al., Eds., *Current Protocols in Molecular Biology* (Wiley, New York, 1994).
16. J. L. Pomerantz, T. M. Kristie, P. A. Sharp, *Genes Dev.* **6**, 2047 (1992).
17. P. E. Pellett, J. L. C. McKnight, F. J. Jenkins, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5870 (1985).
18. Each model of a hybrid protein (Figs. 1 and 5) was constructed by juxtaposition of portions of two different crystallographically determined protein-DNA complexes. Models were initially prepared by superimposition of the double helices in various registers and were analyzed to see how the polypeptide chains might be connected. Superimposition of sets of phosphates typically gave root mean squared distances of 0.5 to 1.5 Å between different sets of [illegible]. These distances [illegible] [illegible]
19. [illegible]
20. The probe used for random binding site selection contained the sequence 5'-GGCTGAGTCTGAACGGATCCN$_{16}$CCTCGAGACTGAGCGTCG-3'. Four rounds of selection were done as described (10), except that 100 ng of poly[d(I-C)]-poly[d(I-C)] and 0.025% NP-40 were included in the binding reactions. Selections used 5 ng of randomized DNA in the first round and approximately 1 ng in subsequent rounds. Binding reactions contained 6.4 ng of GST-ZFHD1 in round 1, 1.6 ng in round 2, 0.4 ng in round 3, and 0.1 ng in round 4.
21. DNA binding reactions contained 10 mM Hepes (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 0.75 mM dithiothreitol, 4% Ficoll-400, and BSA (300 μg/ml), with the appropriate protein and binding site in a total volume of 10 μl. The concentration of binding site was always lower than the apparent dissociation constant by at least a factor of 10. The PA-Oct-1 POU fusion had been described (16). The purified three-finger Zif268 peptide (1) was a gift from M. Elrod-Erickson. Reactions were incubated at 30°C for 30 min and resolved in 4% nondenaturing polyacrylamide gels (16). Apparent dissociation constants were determined as described (10). Probes were derived by cloning of the following fragments into the Kpn I and Xho I sites of pBSKII+ (Stratagene) and excision of the fragments with Asp718 and Hind III: 5'-CCTCGAGGTCATTATGGGCGCTAGGTACC-3', 5'-CCTCGAGGCGCCCATCATTACTAGGTACC-3', 5'-CCTCGAGGCGCCCACGCCTAGGTACC-3', and 5'-CCTCGAGGTCATTTGCATACTAGGTACC-3'.
22. The ZFHD1-VP16 expression vector was constructed by cloning of a fragment encoding the epitope Met-Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala, ZFHD1, and VP16 residues [illegible] into the [illegible] and Aça I sites of Rc/CMV (Invitrogen). Reporter vectors were constructed by cloning of the following fragments into the Xho I and Kpn I sites of pGL2-Promoter (Promega): 5'-GGTACCAGTATGCAAATGACTGCAGTATGCAAATGACCTCGAG-3', 5'-GGTACCAGGCGTGGGCGCTGCAGGCGTGGGCGCCTCGAG-3', and 5'-GGTACCAGTAATGATGGGCGCTGCAGTAATGATGGGCGCCTCGAG-3'. The 293 cells were transfected with the use of calcium phosphate precipitation with a glycerol shock as described (15). Quantitation of human growth hormone (HGH) production was done with the Tandem-R HGH Immunoradiometric Assay (Hybritech, San Diego, CA) according to the manufacturer's instructions. Cell extracts were made 48 hours after transfection (15) and luciferase activity was determined with the use of 10 μl of 100 μl of total extract per 10-cm plate and 100 μl of Luciferase Assay Reagent (Promega) in a ML2250 Luminometer (Dynatech Laboratories, Chantilly, VA) with the use of the enhanced flash program and integration for 20 s with no delay.
23. F. C. M. Ma et al., *Cell* **77**, 451 (1994).
24. E. F. Luisi et al., *Nature* **352**, 497 (1991).

25. We thank J. Crispino, A. MacMillan, and J. Kim for critical reading of the manuscript; D. Shaywitz and J. Klemm for helpful discussions; members of the Sharp lab for their continual support; M. Siafaca for her ever-present assistance; L. Nekludova for help with modeling and programming; W. Xu and R. Sauer for help using Insight; and Y. Qiu and R. Issner for indispensable technical assistance. Supported by USPHS grant PO1-CA42063 from NIH, by cooperative agreement CDR-8803014 from NSF (to P.A.S.), and partially by the National Cancer Institute Cancer Center Support (core) grant P30-CA14051. Graphics were prepared with equipment purchased with support from the PEW Charitable Trusts and the Howard Hughes Medical Institute. C.O.P. is in the Howard Hughes Medical Institute. J.L.P. is partially supported by the Harvard Medical School M.D.-Ph.D. program.

9 September 1994; accepted 31 October 1994

Footprint Analysis of Replicating Murine Leukemia Virus Reverse Transcriptase

Birgitta M. Wöhrl, Millie M. Georgiadis, Alice Telesnitsky, Wayne A. Hendrickson, Stuart F. J. Le Grice*

Replication complexes that contained either murine leukemia virus reverse transcriptase (MLV RT) or a variant reverse transcriptase without a ribonuclease (RNase) H domain (ΔRH MLV RT) were visualized by enzymatic footprinting. Wild-type MLV RT protected template nucleotides +6 to −27, and primer nucleotides −1 to −26 of primers that had first been extended by one or four nucleotides. Although it catalyzed DNA synthesis, ΔRH MLV RT stably bound template-primer only under conditions of reduced ionic strength and protected the duplex portion only as far as position −15. Despite altered hydrolysis profiles, both enzymes covered primarily the template-primer duplex, contradicting recent predictions based on the structure of rat DNA polymerase β.

Even though they catalyze common reactions, retroviral reverse transcriptases (RTs) are structurally diverse. Whereas human, equine, feline, and simian enzymes share a heterodimeric organization of subunits encoded by the RT gene (1–5), a subunit of the avian sarcoma-leukosis virus (ASLV) enzyme retains the integrase domain of the gag-pol polyprotein (6). The isolated enzyme of the murine leukemia virus (MLV) is a 75-kD monomer, whereas a 150-kD homodimer is proposed to catalyze DNA synthesis (7). A structural motif common to RNase H of *Escherichia coli* and MLV RT (α helix III) is also absent from the human immunodeficiency virus (HIV) enzyme (8). Such observations illustrate the importance of comparative studies in understanding the evolution of these multifunctional enzymes.